SEQUENCE LISTING

<110> Gerngross, Tillman U.

<120> METHODS FOR PRODUCING MODIFIED GLYCOPROTEINS

<130> GFI 100

<140> 09/892,591
<141> 2001-06-27

<150> 60/214,358
<151> 2000-06-28

<150> 60/215,638
<151> 2000-06-30

<150> 60/279,997
<151> 2001-03-30

<160> 6

<170> PatentIn version 3.1

<210> 1
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer A for target gene in P. pastoris (1,6-mannosyltransferase)

<400> 1
atggcgaagg cagatggcag t
21

<210> 2
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer B for target gene in P. pastoris (1,6-mannosyltransferase)

.

Table 2.

PCR primer A	PCR primer B	Target Gene(s) in *P. pastoris*	Homologues
ATGGCGAAGGCAGA TGGCAGT (SEQ ID NO:1)	TTAGTCCTTCCAAC TTCCTTC (SEQ ID NO:2)	1,6- mannosyltransferase	OCH1 *S.cerevisiae, Pichia albicans*
TAYTGGMGNGTNGA RCYNGAYATHAA (SEQ ID NO:3)	GCRTCNCCCCANCK YTCRTA (SEQ ID NO:4)	1,2 mannosyltransferases	KTR/KRE family, *S.cerevisiae*

Legend: M = A or C, R = A or G, W = A or T, S = C or G, Y = C or T, K = G or T, V = A or C or G, H = A or C or T, D = A or G or T, B = C or G or T,
N = G or A or T or C.

Incorporation of a Mannosidase into the Genetically Engineered Host

The process described herein enables one to obtain such a structure in high yield for the purpose of modifying it to yield complex N-glycans. A successful scheme to obtain suitable $Man_5GlcNAc_2$ structures must involve two parallel approaches: (1) reducing endogenous mannosyltransferase activity and (2) removing 1,2- α- mannose by mannosidases to yield high levels of suitable $Man_5GlcNAc_2$ structures. What distinguishes this method from the prior art is that it deals directly with those two issues. As the work of Chiba and coworkers demonstrates, one can reduce $Man_8GlcNAc_2$ structures to a $Man_5GlcNAc_2$ isomer in *S. cerevisiae*, by engineering the presence of a fungal mannosidase from *A. saitoi* into the ER. The shortcomings of their approach are twofold: (1) insufficient amounts of $Man_5GlcNAc_2$ are formed in the extra-cellular glycoprotein fraction (10%) and (2) it is not clear that the *in vivo* formed $Man_5GlcNAc_2$ structure in fact is able to accept GlcNAc by action of GlcNAc transferase I. If several glycosylation sites are present in a desired protein the probability (P) of obtaining such a protein in a correct form follows the relationship $P=(F)^n$, where n equals the number of glycosylation sites, and F equals the fraction of desired glycoforms. A glycoprotein with three glycosylation sites would have a 0.1% chance of providing the appropriate

In some cases the library may be assembled directly from existing or wild-type genes. In a preferred embodiment however the DNA library is assembled from the fusion of two or more sub-libraries. By the in-frame ligation of the sub-libraries, it is possible to create a large number of novel genetic constructs encoding useful targeted glycosylation activities. For example, one useful sub-library includes DNA sequences encoding any combination of enzymes such as sialyltransferases, mannosidases, fucosyltransferases, galactosyltransferases, glucosyltransferases, and GlcNAc transferases. Preferably, the enzymes are of human origin, although other mammalian, animal, or fungal enzymes are also useful. In a preferred embodiment, genes are truncated to give fragments encoding the catalytic domains of the enzymes. By removing endogenous targeting sequences, the enzymes may then be redirected and expressed in other cellular loci. The choice of such catalytic domains may be guided by the knowledge of the particular environment in which the catalytic domain is subsequently to be active. For example, if a particular glycosylation enzyme is to be active in the late Golgi, and all known enzymes of the host organism in the late Golgi have a certain pH optimum, then a catalytic domain is chosen which exhibits adequate activity at that pH.

Another useful sub-library includes DNA sequences encoding signal peptides that result in localization of a protein to a particular location within the ER, Golgi, or trans Golgi network. These signal sequences may be selected from the host organism as well as from other related or unrelated organisms. Membrane-bound proteins of the ER or Golgi typically may include, for example, N-terminal sequences encoding a cytosolic tail (ct), a transmembrane domain (tmd), and a stem region (sr). The ct, tmd, and sr sequences are sufficient individually or in combination to anchor proteins to the inner (lumenal) membrane of the organelle. Accordingly, a preferred embodiment of the sub-library of signal sequences includes ct, tmd, and/or sr sequences from these proteins. In some cases it is desirable to provide the sub-library with varying lengths of sr sequence. This may be accomplished by PCR using primers that bind to the 5' end of the DNA encoding the cytosolic region and employing a series of opposing primers that bind to various parts of the stem region. Still other useful sources of signal sequences include retrieval signal peptides, e.g. the tetrapeptides HDEL (SEQ ID NO:5) or KDEL (SEQ ID

NO:6), which are typically found at the C-terminus of proteins that are transported retrograde into the ER or Golgi. Still

other sources of signal sequences include (a) type II membrane proteins, (b) the enzymes listed in Table 3, (c) membrane spanning nucleotide sugar transporters that are localized in the Golgi, and (d) sequences referenced in Table 5.

Table 5. Sources of useful compartmental targeting sequences

Gene or Sequence	**Organism**	**Function**	**Location of Gene Product**
MnsI	*S. cerevisiae*	α-1,2-mannosidase	ER
OCH1	*S. cerevisiae*	1,6-mannosyltransferase	Golgi (cis)
MNN2	*S. cerevisiae*	1,2-mannosyltransferase	Golgi (medial)
MNN1	*S. cerevisiae*	1,3-mannosyltransferase	Golgi (trans)
OCH1	*P. pastoris*	1,6-mannosyltransferase	Golgi (cis)
2,6 ST	*H. sapiens*	2,6-sialyltransferase	trans Golgi network
UDP-Gal T	*S. pombe*	UDP-Gal transporter	Golgi
Mnt1	*S. cerevisiae*	1,2-mannosyltransferase	Golgi (cis)
HDEL (SEQ ID NO:5) at C-terminus	*S. cerevisiae*	retrieval signal	ER

In any case, it is highly preferred that signal sequences are selected which are appropriate for the enzymatic activity or activities which are to be engineered into the host. For example, in developing a modified microorganism capable of terminal sialylation of nascent *N*-glycans, a process which occurs in the late Golgi in humans, it is

desirable to utilize a sub-library of signal sequences derived from late Golgi proteins. Similarly, the trimming of $Man_8GlcNAc_2$ by an α-1,2-mannosidase to give $Man_5GlcNAc_2$